Multivariable Calculus

Thomas Kong

June 29, 2021

Contents

Chapter 1

Preliminaries

1.1 Vector Spaces

Our end goal is to do *multivariable* calculus. That is, we want to differentiate and integrate (real) vector-valued functions of several variables. To this end, we'll need to rigourlously understand what a vector is so that we can comfortably manipulate them (beyond the highschool 'definition' of 'a vector is a quantity with both a magnitude and a direction'). Thus, we'll begin with a lightning review of basic concepts from linear algebra, beginning with the *actual* definition of a vector. A vector is simply an element of a structure called a *vector space*, which we define below:

Definition 1.1.1. A **vector space** over the field $\mathbb{K}$ is a set V, together with the operations of 'vector addition' $f : V \times V \to V$ and 'scalar multiplication' $g : \mathbb{K} \times V \to V$, typically denoted by $f(x, y) = x + y$ and $g(\alpha, x) = \alpha x$, which satisfy the following axioms:

(a) V is an abelian group with respect to the binary operation of vector addition.

(b) Associativity of scalar multiplication: For every $v \in V$ and $a, b \in \mathbb{K}$, $(ab)v = a(bv)$.

(c) Distributivity of scalar multiplication: For every $v \in V$ and $a, b \in \mathbb{K}$, $(a + b)v = av + bv$.

(d) Distributivity of scalar multiplication over vector addition: For every $v, w \in V$ and $a \in \mathbb{K}$, $a(v + w) = av + aw$.

(e) Multiplicative identity: For every $v \in V$, $1v = v$.

We will only consider vector spaces over the field of real numbers; from now on we will let the field $\mathbb{K}$ be $\mathbb{R}$, and herein when we say 'vector space' we are really referring to a *real* vector space. The prototypical example of a vector space is the very first vector space that we all worked with before we even heard the term 'vector space': namely, n-tuples of real numbers, $\mathbb{R}^n$. Indeed, this will basically be the only vector space we will care about, as we will later see.

We'll now fly through some standard linear algebra concepts. Let V be a vector space. A **subspace** is a subset $S \subset V$ if S is itself a vector space under the same vector addition and scalar multiplication operations, restricted to S. Fortunately, to check that a subset of a vector space is a subspace, one does not actually have to check every single one of the vector space axioms. A necessary and sufficient condition for this is simply for the subset to be closed under vector addition and scalar multiplication, i.e for all $x, y \in S$ and $a \in \mathbb{R}$, $x + y \in S$ and $ax \in S$.

A mapping $f : S \to T$ between vector spaces that preserves the vector space structure is called a **linear map** or a **vector space homomorphism**, i.e for all $x, y \in S$ we have $f(x + y) = f(x) + f(y)$ and for all $a \in \mathbb{R}$ we have $f(ax) = af(x)$. A bijective homomorphism is called an **isomorphism**. Two vector spaces are *isomorphic* is there exists an isomorphism between them. Isomorphic vector spaces are essentially 'the same' for all intents and purposes; vector spaces that are isomorphic share all the same properties.

Let $f : S \to T$ be a vector space homomorphism. The set $\ker f = \{x \in S \mid f(x) = 0\}$ is called the **kernel** of f, and the set $\mathcal{R}(f) = \{f(x) \mid x \in S\}$ is called the **range** of f. It can be easily shown that $\ker f$ is a subspace of S and $\mathcal{R}(f)$ is a subspace of T. Another useful fact worth noting is that a linear map f is injective if and only if its kernel is trivial (i.e $\ker f = \{0\}$). Hence, f is an isomorphism if and only if $\ker f = \{0\}$ and $\mathcal{R}(f) = T$.

A particularly important way that we use to classify vector spaces is the notion of dimension, which intuitively speaking is the 'number of degrees of freedom' it possesses. We will proceed to formalise this below. To find out how many degrees of freedom a vector space has, we essentially need to find the minimum number of fixed vectors required to write any arbitrary vector from the space as some weighted sum of these fixed vectors. We call such weighted sums of vectors **linear combinations**; a linear combination of the vectors in the subset A is a sum $\sum_{x \in A} c_x x$. The **span** of a subset A is the set of all possible linear combinations of vectors in A. Hence, in order to describe the entire vector space in terms of sums of vectors from one of its subsets A, we require that $V = \text{span } A$.

However, even if we are able to identify a (proper) subset A of V such that $V = \text{span } A$, it is possible that we are able to find a smaller subset that does the trick, which suggests that A may contain redundant information. The way we describe this redundancy is through the concept of linear dependence:

Definition 1.1.2. Let V be a vector space. A subset $A \subset V$ is **linearly independent** if for every finite subset $\{v_1, \ldots, v_n\}$ of A, we have that $a_1 v_1 + \cdots + a_n v_n = 0$ for scalars $a_1, \ldots, a_n \in \mathbb{K}$ implies that $a_1 = \cdots = a_n = 0$.

If A is not linearly independent, then it is **linearly dependent**.

At last, we arrive at our desired criteria for a subset of a vector space to summarise all of the information of V in the most minimalistic way possible:

Definition 1.1.3. Let V be a vector space. A **basis** for V is a linearly independent subset $A \subset V$ such that $V = \text{span } A$.

Fun remark: The definition we provide above is that of a *Hamel* basis. There are other types of bases. However, this distinction is not relevant for us in the finite dimensional setting, which we will be exclusively working in.

Given a basis $\{v_1, \ldots, v_n\}$ for a vector space V, the representation of a vector $v \in V$ in that basis is unique. That is, if $v = a_1 v_1 + \cdots + a_n v_n$ and $v = b_1 v_1 + \cdots + b_n v_n$ for constants $a_i, b_i \in \mathbb{R}$ for $i = 1, \ldots, n$, we have that $a_i = b_i$ for each i. This fact follows from the linear independence of the basis vectors.

So, we would like to say that the number of elements in a basis quantifies the number of degrees of freedom that the vector space possesses. However, one concern that arises is whether in our definition above that there is a possibility that there are bases with different numbers of elements. Fortunately, the answer is no, as we will now show:

Theorem 1.1.4. *Let V be a vector space. Then every basis of V has the same number of elements, or are all infinite.*

Proof. Suppose $\{v_1, \ldots, v_n\}$ is a basis for V and suppose for a contradiction that $\{x_1, \ldots, x_{n+1}\}$ is a linearly independent subset of V. We can write x_1 as a linear combination of the basis elements:

$$x_1 = a_{11} v_1 + \ldots a_{1n} v_n$$

where $a_{1i} \in \mathbb{R}$ for all $i = 1, \ldots, n$. Since $\{x_1, \ldots, x_{n+1}\}$ is a linearly independent set, then $x_1 \neq 0$, so that not all of the a_{1i}'s are zero. Without loss of generality, suppose that $a_{11} \neq 0$ (else swap its label with one of the a_{1i}'s that is nonzero, and also exchange the indices of the corresponding basis vectors accordingly). Then we can rearrange the above expression to solve for v_1:

$$v_1 = \frac{1}{a_{11}}(x_1 - a_{12} v_2 - \cdots - a_{1n} v_n)$$

It follows that the set $\{x_1, v_2, \dots v_n\}$ is a basis for V. To see this, let $y \in V$. Then $y = c_1 v_1 + \cdots + c_n v_n$ for constants $c_1, \dots, c_n \in \mathbb{R}$ (since the v_i's form a basis for V). Thus, we have that:

$$y = \frac{c_1}{a_{11}} x_1 + \left(c_2 - \frac{a_{12}}{a_{11}}\right) v_2 + \cdots + \left(c_n - \frac{a_{1n}}{a_{11}}\right) v_n$$

So y can be written as a linear combination of the vectors $x_1, v_2, \dots, v_n$, i.e $y \in \operatorname{span}\{x_1, v_2, \dots, v_n\}$. Since the choice of $y \in V$ was arbitrary, we have that $V = \operatorname{span}\{x_1, v_2, \dots, v_n\}$. Linear independence of this set follows from the linear independence of the original basis. Indeed, consider the following vector equation:

$$k_1 x_1 + k_2 v_2 + \cdots + k_n v_n = 0$$

for constants $k_1, \dots, k_n \in \mathbb{R}$. Expanding x_1 in terms of the original basis gives:

$$\begin{aligned} 0 &= k_1(a_{11} v_1 + \dots a_{1n} v_n) + k_2 v_2 + \cdots + k_n v_n \\ &= k_1 a_{11} v_1 + (k_1 a_{12} + k_2) v_2 + \cdots + (k_1 a_{1n} + k_n) v_n \end{aligned}$$

Now, by linear independence of $\{v_1, \dots, v_n\}$, all of the coefficients in the above equation must vanish. In particular, we see that $k_1 a_{11} = 0$, which implies that $k_1 = 0$ since $a_{11} \neq 0$ by assumption. Then, since $k_1 a_{1i} + k_i = 0$ for all $i = 2, \dots, n$, we must have that $k_i = 0$ as well. Hence, $\{x_1, v_2, \dots, v_n\}$ is linearly independent, and our claim is proven.

What have we achieved? We have just replaced one of the basis vectors (namely v_1) with one of the vectors from the linearly independent set (namely x_1), and after the dust cleared we still have a basis for V. We will continue this process, gradually replacing all of the v_i's with an x_i until we have a basis consisting of only x_i's. We will prove that this process will work via induction.

Suppose we have replaced j of the basis vectors, and we have that $\{x_1, \dots, x_j, v_{j+1}, \dots v_n\}$ is a basis for V (after possibly some relabelling of the vectors). We will show that we will obtain a basis by replacing one of the remaining v_i's with x_{j+1}. The argument will follow quite similarly to our first replacement process. Write x_{j+1} in terms of the basis $\{x_1, \dots, x_j, v_{j+1}, \dots v_n\}$:

$$x_{j+1} = a_{j+1,1} x_1 + \dots a_{j+1,j} x_j + a_{j+1,j+1} v_{j+1} + \cdots + a_{j+1,n} v_n$$

Since $x_{j+1} \neq 0$, then not all of the coefficients $a_{j+1,i}$ are zero. In fact, we must have that $a_{j+1,i} \neq 0$ for some $i \geq j+1$ (if this were not the case, then it follows that x_{j+1} is a linear combination of the vectors $x_1, \dots, x_j$, which contradicts the linear independence of the x_i's). Without loss of generality, we'll take $a_{j+1,j+1} \neq 0$. Hence, we can write:

$$v_{j+1} = \frac{1}{a_{j+1,j+1}} (x_{j+1} - a_{j+1,1} x_1 - \cdots - a_{j+1,j} x_j - a_{j+1,j+2} v_{j+2} - \dots a_{j+1,n} v_n)$$

We'll now demonstrate that $\{x_1, \dots, x_{j+1}, v_{j+2}, \dots, v_n\}$ is a basis for V. Let $y \in V$. Writing y in terms of the basis $\{x_1, \dots, x_j, v_{j+1}, \dots v_n\}$ yields:

$$y = c_1 x_1 + \cdots + c_j x_j + c_{j+1} v_{j+1} + \cdots + c_n v_n$$

for some constants $c_1, \dots, c_n \in \mathbb{R}$. Substituting in our expression for v_{j+1} gives:

$$y = \left(c_1 - \frac{a_{j+1,1}}{a_{j+1,j+1}}\right) x_1 + \dots \left(c_j - \frac{a_{j+1,j}}{a_{j+1,j+1}}\right) x_j + \frac{c_{j+1}}{a_{j+1,j+1}} x_{j+1} + \left(c_{j+2} - \frac{a_{j+1,j+2}}{a_{j+1,j+1}}\right) v_{j+2} + \cdots + \left(c_n - \frac{a_{j+1,n}}{a_{j+1,j+1}}\right) v_n$$

Hence $y \in \operatorname{span}\{x_1, \dots, x_{j+1}, v_{j+2}, \dots, v_n\}$. Since the choice of $y \in V$ was arbitrary, it follows that $V = \operatorname{span}\{x_1, \dots, x_{j+1}, v_{j+2}, \dots, v_n\}$.

Now we'll demonstrate linear independence. Consider the following vector equation:

$$k_1 x_1 + \cdots + k_{j+1} x_{j+1} + k_{j+2} v_{j+2} + \cdots + k_n v_n = 0$$

for constants $k_1, \ldots, k_n \in \mathbb{R}$. Expanding x_{j+1} in terms of the basis $\{x_1, \ldots, x_j, v_{j+1}, \ldots v_n\}$ yields:

$$\begin{aligned} 0 &= k_1 x_1 + \cdots + k_{j+1}(a_{j+1,1}x_1 + \ldots a_{j+1,j}x_j + a_{j+1,j+1}v_{j+1} + \cdots + a_{j+1,n}v_n) + k_{j+2}v_{j+2} + \cdots + k_n v_n \\ &= (k_1 + k_{j+1}a_{j+1,1})x_1 + \cdots + (k_j + k_{j+1}a_{j+1,j})x_j + k_{j+1}a_{j+1,j+1}v_{j+1} + (k_{j+2} + k_{j+1}a_{j+1,j+2})v_{j+2} + \\ &\quad \cdots + (k_n + k_{j+1}a_{j+1,n})v_n \end{aligned}$$

By linear independence of $\{x_1, \ldots, x_j, v_{j+1}, \ldots v_n\}$, all of the above coefficients must vanish. In particular, consider the $j+1$ coefficient: $k_{j+1}a_{j+1,j+1} = 0$. From this, we conclude that $k_{j+1} = 0$ since $a_{j+1,j+1} \neq 0$, and hence looking at the remaining coefficients we must have $k_i = 0$ for all $i = 1, \ldots, n$, thus demonstrating the linear independence of $\{x_1, \ldots, x_{j+1}, v_{j+2}, \ldots, v_n\}$, and hence proving that it is a basis for V.

By induction on the finite set $\{1, \ldots, n\}$, we can carry out our replacement procedure and end up with the fact that $\{x_1, \ldots, x_n\}$ forms a basis for V. Now, since $x_{n+1} \in V$, we can write it as a nonzero linear combination of the basis vectors $x_1, \ldots, x_n$. However, this contradicts the linear independence of the set $\{x_1, \ldots, x_{n+1}\}$. We conclude that no linearly independent set can have any more vectors than any basis of V, if there exists a finite basis for V. Since bases are linearly independent sets, then no finite basis can have any more vectors than any other basis. So if there exists a finite basis for V, then every basis for V must also be finite and have the same number of elements. The only other possibility is that every basis of V contains infinitely many elements. □

This justifies the following definition:

Definition 1.1.5. Let V be a vector space. The **dimension** of V, denoted $\dim V$, is the number of elements in any basis of V if they are finite, or ∞ otherwise.

We will only consider finite dimensional vector spaces, that is, vector spaces in which there exists a basis with finitely many elements. Doing calculus on infinite dimensional vector spaces crosses into the realm of functional analysis, but we won't stray in that direction in the scope of these notes.

One key advantage of working in finite dimensional spaces is that things all turn out to be significantly simpler. One easy consequence of the previous theorem is that in an n-dimensional vector space, if we have a linearly independent set consisting of n elements, it must actually span the entire space and is thus automatically a basis.

Corollary 1.1.6. *Let V be an n-dimensional vector space and let $\{v_1, \ldots, v_n\}$ be a linearly independent set. Then $\{v_1, \ldots, v_n\}$ is a basis for V.*

Proof. Suppose for a contradiction that $\operatorname{span}\{v_1, \ldots, v_n\} \neq V$. So there exists a $x \in V$ such that x is not a linear combination of the v_i's. Hence, $\{v_1, \ldots, v_n, x\}$ is a linearly independent set. So we have a linearly independent set consisting of $n+1$ elements in an n-dimensional space. This is impossible, since every basis of V must contain n elements and no linearly independent set can have more elements than any basis of V. We must conclude that $V = \operatorname{span}\{v_1, \ldots, v_n\}$, and hence $\{v_1, \ldots, v_n\}$ is a basis for V. □

Another simplification in the finite dimensional setting is that there actually aren't 'that many' different types of finite dimensional vector spaces for each dimension $n \in \mathbb{N}$ to study; we can easily completely classify every real finite dimensional vector space up to isomorphism - even better: they are all structurally identical to $\mathbb{R}^n$!

Theorem 1.1.7. *Let V be an n-dimensional vector space. Then V is isomorphic to $\mathbb{R}^n$.*

Proof. Let $\{v_1, \ldots, v_n\}$ be a basis for V. Define the mapping $\phi : V \to \mathbb{R}^n$ to act on $v = a_1 v_1 + \cdots + a_n v_n$ by:

$$\begin{aligned} \phi(v) &= \phi(a_1 v_1 + \cdots + a_n v_n) \\ &= (a_1, \ldots, a_n)^T \end{aligned}$$

So ϕ is the canonical map that maps a vector $v \in V$ to an n-tuple of real numbers, which are simply the coordinates of v with respect to a certain basis of V. We will show that ϕ is an isomorphism.

Note that the elements mapped by ϕ to $0 \in \mathbb{R}^n$ must have zero coefficients when expanded in terms of the basis $\{v_1, \ldots, v_n\}$, and thus can only be the zero vector. This shows that $\ker \phi = \{0\}$ and hence that ϕ is injective.

Let $x \in \mathbb{R}^n$. Then $x = (a_1, \ldots, a_n)^T$ for some $a_1, \ldots, a_n \in \mathbb{R}$. Thus, it is easy to see that the element $v = a_1 v_1 + \cdots + a_n v_n \in V$ is mapped to x by ϕ. Since the choice of $x \in \mathbb{R}^n$ was arbitrary, it follows that ϕ is surjective.

Let $v = a_1 v_1 + \cdots + a_n v_n \in V$ and $w = b_1 v_1 + \cdots + b_n v_n \in V$. Then:

$$\begin{aligned} \phi(v + w) &= (a_1 + b_1, \ldots, a_n + b_n) \\ &= (a_1, \ldots, a_n) + (b_1, \ldots, b_n) \\ &= \phi(v) + \phi(w) \end{aligned}$$

Let $v \in V$ be as above and let $k \in \mathbb{R}$. Then:

$$\begin{aligned} \phi(kv) &= (ka_1, \ldots, ka_n) \\ &= k(a_1, \ldots, a_n) \\ &= k\phi(v) \end{aligned}$$

Thus, ϕ is a vector space homomorphism. Together with the fact that ϕ is bijective, we have hence shown that ϕ is an isomorphism. $\square$

So any two n-dimensional vector spaces are isomorphic. This is what we meant when we said that the only n-dimensional vector space we will care about is $\mathbb{R}^n$ - all other instances of an n-dimensional vector space are algebraically equivalent, so we may as well just consider this simple example as being *the* n-dimensional vector space.

1.2 Euclidean Space

Now that we have formulated all of the algebraic structure of the multidimensional spaces we will be working in, we will need to endow some further structure onto these spaces in order to do calculus. Of great importance is the ability to measure distances between points/vectors in our space. At the heart of analysis/calculus is the notion of *limits* and *convergence*. We will also frequently want to find bounds and estimates for quantities that may very well be vector-valued. So what we ultimately want is to prescribe some notion of 'length' for vectors, which will in turn naturally define a distance between two vectors (simply take the length of the difference of the two vectors). This is captured by the concept of a vector *norm*.

Definition 1.2.1. Let V be a vector space. A norm on V is a function $\|-\| : V \to \mathbb{R}$ satisfying:

(a) Positivity: For all $x \in V$ $\|x\| \geq 0$. Furthermore, $\|x\| = 0$ if and only if $x = 0$.

(b) Homogeneity of degree 1: For all $x \in V$ and $a \in \mathbb{R}$, $\|ax\| = |a|\|x\|$.

(c) Triangle inequality: for all $x, y \in V$, $\|x + y\| \leq \|x\| + \|y\|$.

A vector space together with a norm is called a **normed vector space**.

Another useful concept from coordinate geometry is the notion of angles. This will lend us the ability to define properties such as orthogonality, and perform operations such as projecting a vector onto a subspace. This structure is given rise to by an *inner product*, which we define below:

Definition 1.2.2. Let V be a vector space. A (real) **inner product** on V is a function $\langle -, - \rangle : V \times V \to \mathbb{R}$ satisfying:

(a) Symmetry: for all $x, y \in V$, $\langle x, y \rangle = \langle y, x \rangle$.

(b) Distributivity: for all $x, y, z \in V$, $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.

(c) Homogeneity of degree 1: For all $x, y \in V$ and $a \in \mathbb{R}$, $\langle ax, y\rangle = a\langle x, y\rangle$.

(d) Positivity: For all $x \in V$, $\langle x, x\rangle \geq 0$. Moreover, $\langle x, x\rangle = 0$ if and only if $x = 0$.

A vector space together with an inner product is called an **inner product space**.

By symmetry of the inner product, we can deduce from the definition that $\langle x, ay\rangle = a\langle x, y\rangle$ and $\langle x, y+z\rangle = \langle x, z\rangle + \langle y, z\rangle$. Hence, an inner product on a vector space is simply a bilinear function (i.e it becomes a linear function of one variable if we fix the value of the other input variable).

You are likely already quite familiar with an inner product on $\mathbb{R}^n$ from coordinate geometry, more commonly known as the dot product or scalar product. Given elements $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$, we define the dot product as:

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$$

The notation $x \cdot y$ is more commonly used instead of $\langle x, y\rangle$ in this context. It is not too hard to see that the dot product is indeed an inner product on $\mathbb{R}^n$.

One powerful fact about inner products is that they can be used to define a norm in a very natural way. Let V be an inner product space. Then for any $x \in V$, we set the norm of x to be given by $\|x\| = \langle x, x\rangle^{\frac{1}{2}}$. We say that the inner product 'induces' a norm on V. So in some sense, once we had defined the notion of angles on our vector space, we have already implicitly fixed some compatible notion of distance as well for free. The positivity and homogeneity properties of the norm follow readily from the properties of the inner product and from the definition of the induced norm. Proving the triangle inequality will take a little more work, and will rely on an identity known as the *Schwarz inequality*:

Theorem 1.2.3. *(Schwarz Inequality) Let V be an inner product space. For every $x, y \in V$, we have:*

$$\langle x, y\rangle \leq \|x\|\|y\|$$

Proof. Let $x, y \in V$ and let $t \in \mathbb{R}$. By positivity of the inner product, we have that $\langle x - ty, x - ty\rangle \geq 0$. This can be rewritten as:

$$\begin{aligned} 0 &\leq \langle x - ty, x - ty\rangle \\ &= \langle x, x\rangle - t\langle x, y\rangle - t\langle y, x\rangle + t^2\langle y, y\rangle \\ &= \|x\|^2 - 2t\langle x, y\rangle + t^2\|y\|^2 \end{aligned}$$

So we have a real quadratic equation in the variable t. The quadratic can have at most a single zero, since $\langle x - ty, x - ty\rangle = 0$ if and only if $x - ty = 0$, i.e $x = ty$. Hence, the discriminant of the quadratic must be either 0 (which corresponds to the single zero) or negative (no real zeros). The discriminant of the quadratic is:

$$4\langle x, y\rangle^2 - 4\|x\|^2\|y\|^2$$

Putting this all together, we conclude that:

$$\begin{aligned} 4\langle x, y\rangle^2 - 4\|x\|^2\|y\|^2 &\leq 0 \\ \langle x, y\rangle &\leq \|x\|\|y\| \end{aligned}$$

□

From the above proof, we note that equality occurs in the Schwarz inequality if and only if $x = ty$ (i.e the vectors are scalar multiples of each other).

We are now ready to tackle the proof that the inner product gives rise to a vector norm:

Theorem 1.2.4. *Let V be an inner product space. The function $\|-\| : V \to \mathbb{R}$ defined by $\|x\| = \langle x, x\rangle^{\frac{1}{2}}$ is a norm on V.*

Proof. Positivity: Let $x \in V$. We have that $\|x\|^2 = \langle x, x \rangle \geq 0$, and hence $\|x\| \geq 0$. Now, from $\|x\|^2 = \langle x, x \rangle$, we can easily see that $\|x\| = 0$ if and only if $x = 0$, by positivity of the inner product.

Homoegenity of degree 1: Let $x \in V$ and $a \in \mathbb{R}$. Then $\|ax\|^2 = \langle ax, ax \rangle = a^2 \langle x, x \rangle = a^2\|x\|^2$. Taking square roots gives $\|ax\| = |a|\|x\|$.

Triangle inequality: Let $x, y \in V$. Then:

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y \rangle \\
&= \langle x, x \rangle + 2\langle x, y \rangle + \langle y, y \rangle \\
&= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \\
&\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 && \text{(by the Schwarz inequality)} \\
&= (\|x\| + \|y\|)^2
\end{aligned}$$

Taking square roots gives the desired result: $\|x+y\| \leq \|x\| + \|y\|$. □

Finally, we need to formalise the notion of measuring distance in our vector space. The idea of distance is captured by a 'metric', which intuitively speaking is a distance-measuring function.

Definition 1.2.5. Let X be a set. A **metric** on X is a function $d : X \times X \to \mathbb{R}$ satisfying:

(a) Positivity: for all $x, y \in X$, $d(x, y) \geq 0$. Moreover, $d(x, y) = 0$ if and only if $x = y$.

(b) Symmetry: for all $x, y \in X$, $d(x, y) = d(y, x)$.

(c) Triangle inequality: for all $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

A set together with a metric is called a **metric space**.

We mentioned earlier that once we had a norm on a vector space, this was sufficient to define a notion of distance. Specifically, we claimed that to measure the distance between two vectors x and y, we simply need to compute the length of the vector $x - y$ using the vector norm. Thus, we claim that the norm naturally induces a metric on the vector space.

Theorem 1.2.6. *Let V be a normed vector space. The function $d : V \times V \to \mathbb{R}$ defined by $d(x, y) = \|x - y\|$ is a metric on V.*

Proof. Positivity: Let $x, y \in V$. Then $d(x, y) = \|x - y\| \geq 0$ by positivity of the norm. Furthermore, $d(x, y) = 0$ if and only if $x - y = 0$, i.e $x = y$.

Symmetry Let $x, y \in V$. Then $d(x, y) = \|x - y\| = \|-(y - x)\| = |-1|\|y - x\| = d(y, x)$, which follows from the homogeneity of the norm.

Triangle inequality: Let $x, y, z \in V$. Then:

$$\begin{aligned}
d(x, z) &= \|x - z\| \\
&= \|(x - y) + (y - z)\| \\
&\leq \|x - y\| + \|y - z\| && \text{(since the norm obeys the triangle inequality)} \\
&= d(x, y) + d(y, z)
\end{aligned}$$

□

In summary, by defining an inner product on our vector space, the inner product induces a norm, which in turn induces a metric. So we get the geometric concepts of angle, length and distance all at once via an inner product. Hence, we call an n-dimensional vector space with an inner product **Euclidean n-space**; where 'Euclidean' refers to the fact that we have a means to measure distances in our space.

Returning to the example of $\mathbb{R}^n$, we see that $\mathbb{R}^n$ together with the dot product defines a Euclidean n-space. In fact, it is interesting to note that the metric induced by the dot product is simply the function that returns the Euclidean distance between two points in $\mathbb{R}^n$ (i.e the Euclidean metric d_2):

$$d_2(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

In the previous section, we showed that all n-dimensional vector spaces are isomorphic to each other, and in particular, $\mathbb{R}^n$, so we can just talk about $\mathbb{R}^n$ without any loss of generality. Even better, it is the case that all n-dimensional inner product spaces are isomorphic, and moreover the inner product is preserved by the isomorphism. Thus, there is essentially only one Euclidean n-space up to isomorphism: namely, $\mathbb{R}^n$ with the dot product. The proof of this will be the subject of the following section.

1.3 Orthonormal Basis

Most of the time when we are working with vector spaces, it is very useful to pick a basis and represent all vectors in terms of that basis. Some bases are more convenient to work with than others. Indeed, in a Euclidean space, now that we have some notion of angles provided by the inner product, we can define what it means for two vectors to be orthogonal (the generalisation of perpendicularity from coordinate geometry). Hence, we could try to find bases that are *orthonormal*; the basis vectors are orthogonal and normalised. We'll define these concepts below:

Definition 1.3.1. Let E be a Euclidean n-space. Two elements $x, y \in E$ are **orthogonal** if $\langle x, y\rangle = 0$.

A subset $A \subset E$ is said to be orthogonal if every pair of distinct elements from A are orthogonal.

Definition 1.3.2. Let E be a Euclidean n-space. A set $A \subset E$ is **orthonormal** if A is orthogonal and $\|x\| = 1$ for all $x \in A$.

Orthonormality is a powerful condition. One immediate consequence of the orthonormality of a set is linear independence:

Proposition 1.3.3. *Every orthonormal set in a Euclidean n-space is linearly independent.*

Proof. Suppose $\{e_1, \ldots, e_m\}$ is an orthonormal set in a Euclidean n-space E $(m \le n)$. Consider the vector equation $a_1e_1 + \cdots + a_me_m = 0$ for constants $a_1, \ldots, a_n \in \mathbb{R}$. Taking the inner product of e_i, $i = 1, \ldots, m$ with the above equation yields:

$$\begin{aligned} 0 &= \left\langle e_i, \sum_{j=1}^{m} a_j e_j \right\rangle \\ &= \sum_{j=1}^{m} a_j \langle e_i, e_j\rangle \\ &= \sum_{j=1}^{m} a_j \delta_{ij} \\ &= a_i \end{aligned}$$

where we have made use of the Kronecker delta symbol δ_{ij}, which returns 1 if $i = j$ and 0 otherwise.

So $a_1 = 0$ for all $i = 1, \ldots, m$. This proves the linear independence of the finite set $\{e_1, \ldots, e_m\}$.

Now suppose for a contradiction there exists an orthonormal set $A \subset E$ with more than n elements. Pick n elements from A to form the orthonormal set $\{e_1, \ldots, e_n\}$, which is linearly independent from what we have just proven. Hence, it forms a basis for E. Pick another *distinct* element $e_{n+1} \in A$, and write it as a linear combination of the basis elements:

$$e_{n+1} = k_1 e_1 + \cdots + k_n e_n$$

Now take the inner product of the above equation with e_{n+1}:

$$\begin{aligned} \langle e_{n+1}, e_{n+1} \rangle &= \left\langle e_{n+1}, \sum_{j=1}^{n} k_j e_j \right\rangle \\ &= \sum_{j=1}^{n} k_j \langle e_{n+1}, e_j \rangle \\ &= 0 \end{aligned}$$

However, $\langle e_{n+1}, e_{n+1} \rangle = \|e_{n+1}\|^2 = 1$, so we have that $1 = 0$; a contradiction. It follows that there cannot exist an orthonormal set in E with more than n elements. Since we have proven that every orthonormal set with $m \leq n$ elements is linearly independent, then we thus conclude that every orthonormal set in E is linearly independent. □

One interesting consequence of the above proof is the following. Suppose we could write a vector x from a Euclidean space E as a linear combination of orthonormal vectors $x = a_1 e_1 + \cdots + a_n e_n$. Then we can define a mapping $\pi_i : E \to \mathbb{R}$ for $i = 1, \ldots, n$ by $\pi_i(x) = \langle x, e_i \rangle = a_i$. This illustrates another advantage of using an orthonormal basis to represent vectors: finding the coefficients becomes a straightforward task - one simply needs to calculate $\pi_i(x)$ to find the i^{th} coefficient. We call the mapping π_i a *projection operator*; essentially π_i is projecting the vector x onto the subspace spanned by the i^{th} basis vector and picking out the amplitude of the vector projection.

Now that we are sufficiently hyped up about orthonormal bases, the next natural question to ask is whether it is always possible to find an orthonormal basis for any Euclidean n-space. We know from the previous result that every orthonormal set is linearly independent. Combined with the fact that in an n-dimensional space, any linearly independent set consisting of n elements is automatically a basis, all we need to do is prove the existence of an orthonormal set consisting of n elements. Spoiler alert: yes, this is certainly possible. Better yet, the proof of this result will be constructive, it illustrates a standard algorithm called the 'Gram-Schmidt' process which transforms an arbitrary basis into an orthonormal basis.

Proposition 1.3.4. *Let E be a Euclidean n-space. Then there exists an orthonormal subset $A \subset E$ consisting of n elements.*

Proof. Let $\{v_1, \ldots, v_n\}$ be a linearly independent set in E. It is worth pointing out that none of the vectors are 0 due to linear independence. Thus, we can create an orthonormal set of one element by taking the first element from the set and normalising it: define $e_1 = v_1/\|v_1\|$. Then $\{e_1\}$ is an orthonormal set, and e.g $\{e_1, v_2\}$ is a linearly independent set (since e_1 is simply a scalar multiple of v_1). This sets us up to prove the inductive step.

Suppose that we had that $\{e_1, \ldots, e_k\}$ is an orthonormal set ($k < n$) and that $\{e_1, \ldots, e_k, v_{k+1}\}$ is a linearly independent set. Define:

$$y_{k+1} = v_{k+1} - \sum_{j=1}^{k} \langle v_{k+1}, e_j \rangle e_j$$

It follows that $y_{k+1} \neq 0$ since y_{k+1} is a linear combination of $e_1, \ldots, e_k, v_{k+1}$, which is a linearly independent set - and so the only way for such a linear combination to yield the zero vector is if all of the coefficients are 0, which is not the case here. Hence, $\|y_{k+1}\| \neq 0$, and so we can safely define:

$$e_{k+1} = \frac{y_{k+1}}{\|y_{k+1}\|}$$

So $\|e_{k+1}\| = 1$, and for all $i = 1, \ldots, k$:

$$\begin{aligned}
\langle e_{k+1}, e_i \rangle &= \left\langle \frac{1}{\|y_{k+1}\|} \left(v_{k+1} - \sum_{j=1}^{k} \langle v_{k+1}, e_j \rangle e_j \right), e_i \right\rangle \\
&= \frac{1}{\|y_{k+1}\|} \left(\langle v_{k+1}, e_i \rangle - \sum_{j=1}^{k} \langle v_{k+1}, e_i \rangle \langle e_j, e_i \rangle \right) \\
&= \frac{1}{\|y_{k+1}\|} \left(\langle v_{k+1}, e_i \rangle - \sum_{j=1}^{k} \langle v_{k+1}, e_j \rangle \delta_{ij} \right) \\
&= \frac{1}{\|y_{k+1}\|} (\langle v_{k+1}, e_i \rangle - \langle v_{k+1}, e_i \rangle) \\
&= 0
\end{aligned}$$

Hence, $\{e_1, \ldots, e_{k+1}\}$ is an orthonormal set. If $k+1 < n$, then there is still an element v_{k+2} that cannot be written as a linear combination of the orthonormal set $\{e_1, \ldots, e_{k+1}\}$, since the e_i's are simply linear combinations of $v_1, \ldots, v_{k+1}$, but $\{v_1, \ldots, v_{k+2}\}$ is a linearly independent set. It follows that $\{e_1, \ldots, e_{k+1}, v_{k+2}\}$ is a linearly independent set. We can then proceed via induction to deduce that we can produce an orthonormal set $\{e_1, \ldots, e_n\}$. □

The previous two results together imply the following:

Theorem 1.3.5. *Every Euclidean n-space has an orthonormal basis.*

Let us return to the original problem that we wished to address in this section: that there is only one Euclidean n-space up to isomorphism. Since we have already proven that the vector space structure of two Euclidean n-spaces are equivalent, we simply need to prove that the 'extra structure' of the Euclidean space is also preserved under a vector space isomorphism. Let's first make explicit what 'extra stuff' needs to be preserved:

Definition 1.3.6. Let E and F be Euclidean spaces. We say that E and F are **isomorphic** if there exists a mapping $\phi : E \to F$ such that ϕ is a vector space isomorphism that also preserves the inner product, i.e for all $x, y \in E$, $\langle x, y \rangle = \langle \phi(x), \phi(y) \rangle$.

So we require that just the inner product needs to be preserved by the isomorphism. But what about the corresponding norm and metric? Well, if the inner product between Euclidean spaces E and F is preserved by an isomorphism $\phi : E \to F$, it follows that ϕ also preserves the induced norms and metrics on these spaces. To see this, let $x \in E$ and notice that $\|x\|^2 = \langle x, x \rangle = \langle \phi(x), \phi(x) \rangle = \|\phi(x)\|^2$. Now that we have shown that ϕ is norm-preserving, the fact that distances are preserved follows easily. For any $x, y \in E$ we have that $d(x, y) = \|x - y\| = \|\phi(x - y)\| = \|\phi(x) - \phi(y)\| = d(\phi(x), \phi(y))$. Hence, we see that ϕ is an isometry (a distance preserving function).

Recall from our proof that every n-dimensional vector space is isomorphic (to $\mathbb{R}^n$) (Theorem 1.1.7), we constructed a function that essentially mapped one basis onto another. In order to preserve the inner product of the Euclidean spaces, the isomorphisms that we will construct in the following proof will be a subset of the isomorphisms that did the trick in the more general vector space case. Specifically, the isomorphisms between Euclidean spaces will, loosely speaking, send *orthonormal bases to orthonormal bases.*

Theorem 1.3.7. *Let E and F be Euclidean n-spaces. Then E and F are isomorphic.*

Proof. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis in E, and $\{\overline{e_1}, \ldots, \overline{e_n}\}$ an orthonormal basis in F (we know that orthonormal bases for Euclidean spaces exist by the previous theorem which we have just proven). We can express any $x \in E$ as a linear combination of the basis vectors: $x = a_1 e_1 + \cdots + a_n e_n$. We can hence define the map $\phi : E \to F$ by:

$$\phi(x) = a_1 \overline{e_1} + \cdots + a_n \overline{e_n}$$

The form of ϕ is exactly that of the isomorphism which we have constructed in the proof of Theorem 1.1.7, i.e a map that sends a vector represented in a given basis to a vector in the other space with the same coordinates/coefficients but with respect to a basis in the target space[1]. Thus, we already have that ϕ is a vector space isomorphism. We thus simply need to demonstrate that ϕ preserves the inner product. Let $x \in E$ as above and let $y = b_1e_1 + \cdots + b_ne_n \in E$. Then:

$$\begin{aligned}
\langle x, y\rangle &= \left\langle \sum_{i=1}^{n} a_i e_i, \sum_{j=1}^{n} b_j e_j \right\rangle \\
&= \sum_{i=1}^{n} a_i \left\langle e_i, \sum_{j=1}^{n} b_j e_j \right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \langle e_i, e_j\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \delta_{ij} && \text{(by orthonormality of the basis vectors)} \\
&= \sum_{i=1}^{n} a_i b_i
\end{aligned}$$

But we also have that:

$$\begin{aligned}
\langle \phi(x), \phi(y)\rangle &= \left\langle \sum_{i=1}^{n} a_i \overline{e_i}, \sum_{j=1}^{n} b_j \overline{e_j} \right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \langle \overline{e_i}, \overline{e_j}\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \delta_{ij} && \text{(by orthonormality of the basis vectors)} \\
&= \sum_{i=1}^{n} a_i b_i
\end{aligned}$$

So $\langle x, y\rangle = \langle \phi(x), \phi(y)\rangle$, and so the inner product is preserved. Hence, ϕ is an isomorphism of Euclidean spaces. $\square$

Hence, every Euclidean n-space is 'the same' from a structural point of view, so we might as well work with the easiest example of it we can think of: i.e $\mathbb{R}^n$ with the dot product.

While on the topic of orthonormal bases, we'll conclude this section with a fun digression. Let E be a Euclidean n-space, and suppose we had two orthonormal bases in E: $\{e_1, \dots, e_n\}$ and $\{\overline{e_1}, \dots, \overline{e_n}\}$. We can write the barred basis vectors $\overline{e_i}$ in terms of the basis vectors e_j:

$$\overline{e_i} = \sum_{j=1}^{n} a_{ij} e_j \quad i = 1, \dots, n$$

The n^2 constants $a_{ij} \in \mathbb{R}$ form a $n \times n$ matrix O, which represents the linear map $\phi : E \to E$ defined by $\phi(a_1e_1 + \cdots + a_ne_n) = a_1\overline{e_1} + \cdots + a_n\overline{e_n}$, which we know from above to be an automorphism of Euclidean spaces. So if we wrote the components of a vector $x \in E$ with respect to the orthonormal basis $\{e_1, \dots, e_n\}$ as a column vector, then $\phi(x) = Ox$, where the components of $\phi(x)$ are also expressed in a column vector. What properties does the matrix O have? Well, since $\langle \overline{e_i}, \overline{e_j}\rangle = \delta_{ij}$ by orthonormality, we can substitute our

[1] Specifically, we mapped a basis from an n-dimensional vector space to the standard basis in $\mathbb{R}^n$ in that proof.

expression for the barred basis vectors in terms of the unbarred basis vectors to obtain:

$$\begin{aligned}
\delta_{ij} &= \left\langle \sum_{k=1}^{n} a_{ik} e_k, \sum_{l=1}^{n} a_{jl} e_l \right\rangle \\
&= \sum_{k=1}^{n} \sum_{l=1}^{n} a_{ik} a_{jl} \langle e_k, e_l \rangle \\
&= \sum_{k=1}^{n} \sum_{l=1}^{n} a_{ik} a_{jl} \delta_{kl} \\
&= \sum_{k=1}^{n} a_{ik} a_{jk}
\end{aligned}$$

Suppose that b_{ij} are the components of the matrix O^T (the matrix transpose of O). Then $b_{kj} = a_{jk}$, so that we have that $\delta_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. Since δ_{ij} are the components of the $n \times n$ identity matrix I_n, we can write the above equation in the following matrix form:

$$I_n = OO^T$$

Since O is the matrix representation of a vector space automorphism, it is invertible, and hence by premultiplying both sides of the above equation by O^{-1}, we arrive at:

$$O^{-1} = O^T$$

i.e the matrix O is an orthogonal matrix. This reveals that orthogonal transformations between Euclidean spaces are simply the ones that map orthonormal bases to orthonormal bases.

1.4 Dual Space

Definition 1.4.1. Let V be an n-dimensional vector space. A **linear functional** on V is a linear mapping $f : V \to \mathbb{R}$.

Definition 1.4.2. Let V be an n-dimensional vector space. The **dual space** V^* is the set of all linear functionals on V.

The dual space of a vector space is itself a vector space under the operations of addition and scalar multiplication of real-valued functions.

Proposition 1.4.3. *Let V be an n-dimensional vector space and let $\{v_1, \ldots, v_n\}$ be a basis for V. Then:*

(a) Every linear functional $f \in V^*$ is completely determined by its values at $v_1, \ldots, v_n$.

(b) For any real numbers $a_1, \ldots, a_n \in \mathbb{R}$ there exists an $f \in V^*$ such that $f(v_i) = a_i$ for all $i = 1, \ldots, n$.

Theorem 1.4.4. *Let V be an n-dimensional vector space. Then its dual space V^* is also an n-dimensional vector space.*

Theorem 1.4.5. *Let V be an n-dimensional vector space. Then V is isomorphic to its second dual V^{**}.*

Proposition 1.4.6. *Let V be an n-dimensional normed vector space. Then V^* is a normed space, with norm given by:*

$$\|f\| = \sup \{\|f(x)\| \mid x \in V, \|x\| = 1\}$$

Theorem 1.4.7. *Let E be a Euclidean n-space. Then its dual space E^* is also a Euclidean n-space.*

1.5 The Space $L(E,F)$

1.6 Topology of Euclidean Space

1.7 Completeness

1.8 Compactness

1.9 Equivalence of Norms